We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

17

Data Mining in Web Applications

Julio Ponce[1], Alberto Hernández[2], Alberto Ochoa[4,5], Felipe Padilla[3], Alejandro Padilla[1], Francisco Álvarez[1] and Eunice Ponce de León[1]
[1]*Aguascalientes University,*
[2]*CIICAp-UAEM,*
[3]*UQAM,*
[4]*Juarez City University*
[5]*CIATEC*
[1,2,4,5]*México*
[3]*Canada*

1. Introduction

The World Wide Web is rapidly emerging as an important medium for commerce as well as for the dissemination of information related to a wide range of topics (e.g., business and government). According to most predictions, the majority of human information will be available on the Web. These huge amounts of data raise a grand challenge, namely, how to turn the Web into a more useful information utility (Garofalakis et al., 1999) .
At the moment with the popularity of Internet, people are exhibited to a lot of information that is available for study. Nowadays there is also a great amount of applications and services that are available through Internet as they are seeking, chats, sales, etc., nevertheless much of that information is not useful for many people, but in the area of Data Mining, all the information available in the Internet represents a work opportunity and it is possible to do a lot of analysis on the basis of these with specific purposes.
Knowledge Discovery and Data Mining are powerful data analysis tools. The rapid dissemination of these technologies calls for an urgent examination of their social impact. We show an overview of these technologies. The terms "Knowledge Discovery" and "Data Mining" are used to describe the 'non-trivial extraction of implicit, previously unknown and potentially useful information from data (Wahlstrom & Roddick, 2000). Knowledge discovery is a concept that describes the process of searching on large volumes of data for patterns that can be considered knowledge about the data. The most well-known branch of knowledge discovery is data mining.

1.1 Data mining

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential. Data mining is a knowledge discovery process in large and complex data sets, refers to extracting or "mining" knowledge from large amounts of data. Moreover, data mining can be used to predict an outcome for a given entity (Hernández et al., 2006).
Thus clustering algorithms in data mining are equivalent to the task of identifying groups of records that are similar between themselves but different from the rest. (Varan, 2006).

Source: Data Mining and Knowledge Discovery in Real Life Applications, Book edited by: Julio Ponce and Adem Karahoca,
ISBN 978-3-902613-53-0, pp. 438, February 2009, I-Tech, Vienna, Austria

Data mining is a multidisciplinary field with many techniques. With this techniques you can create a mining model that described the data that you will use. (see Fig. 1).

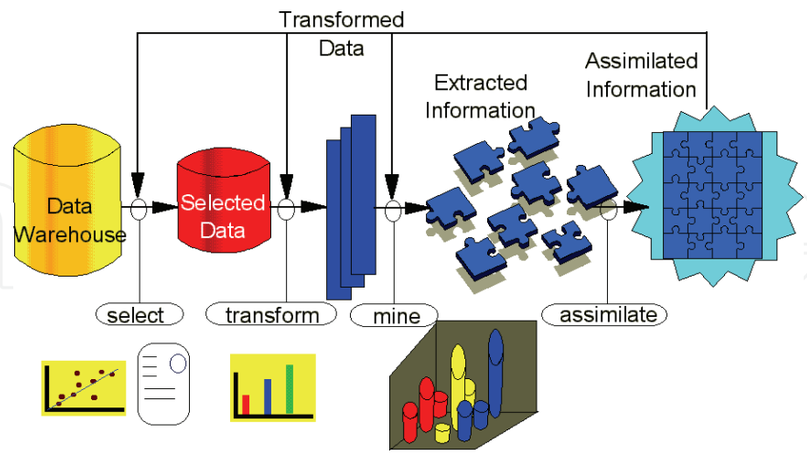

Fig. 1. Data Mining Process

Some elements in Data Mining Process are:

A. **Data Set.** It is a collection of data, usually presented in tabular form. Each column represents a particular variable. Each row corresponds to a given member of the data set in question. It lists values for each of the variables, such as height and weight of an object or values of random numbers. Each value is known as a datum. The data set may comprise data for one or more members, corresponding to the number of rows.
B. **Pre-processing.** Data mining requires substantial pre-processing of data. This was especially the case of the behavioural data. To make the data comparable, all data needs to be normalized.
C. **General Results.** This activity is related to overall assessment of the effort in order to find out whether some important issues might have been overlooked. This is the step where a decision upon further steps has to be made. If all previous steps were satisfactory and results fulfil problem objectives, the project can move to its conclusive phase.
D. **Decision Trees.** Decision trees are powerful and popular tools for classification and prediction. The attractiveness of decision trees is due to the fact that, these decision trees represent rules.Decision tree induction is a typical inductive approach to learn knowledge on classification. The key requirements to do mining with decision trees are: Attribute-value description, Predefined classes, Discrete classes and Sufficient data.
E. **Association Rules.** Association rules describe events that tend to occur together. They are formal statements in the form of X=>Y, where if X happens, Y is likely to happen (Márquez et al., 2008).

1.2 Weka

Weka (Waikato Environment for Knowledge Analysis) is a collection of algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java program. This contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization (Witten & Frank 2005). Weka was developed at the University of Waikato in New Zealand, is an open source software issued under the GNU General Public Lisense. The Data Mining process with Weka includes: reading the

.arrf file in the Weka Explorer, proceeding to classify, visualize clusters and discover associations in the data. Start the hidden patterns finding, remember to keep mind open (No prejudices).

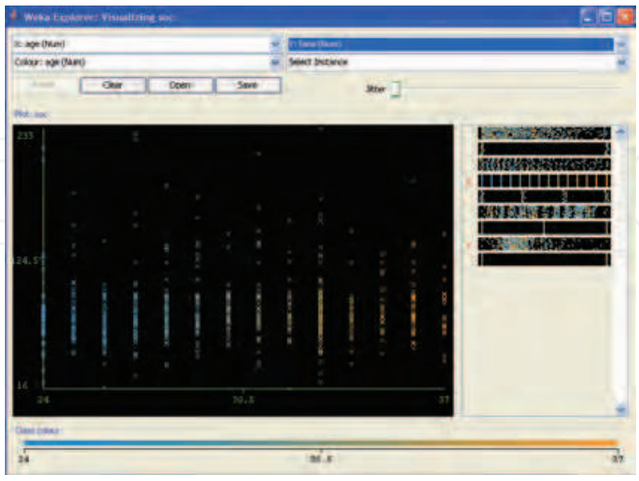

Fig. 2. Visualizing some data in Weka Explores

1.3 Web mining

Usually, web mining is categorized as web content mining and web usage mining. The first studies the search and retrieval of information on the web, while the second discovers and analyzes user's access pattern (Xu et al., 2003). A knowledge discovery tool, WebLogMiner, is discussed in (Zaiane et al., 1998), which uses OLAP and data mining techniques for mining web server log files. In (Mobasher et al, 2000), a web mining frame-work which integrated both usage and content attributes of a site is described. Some techniques based on clustering and association rules are proposed. In (Yao & Yao, 2003; Yao, 2003) presents a framework and information retrieval techniques to support individual scientists doing research.

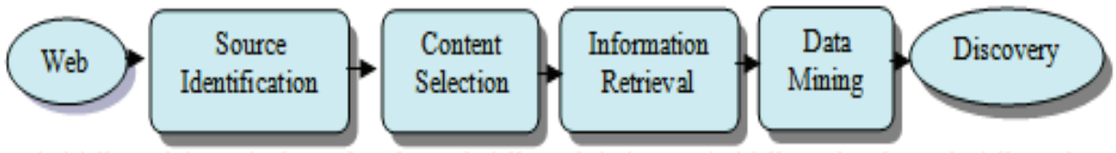

Fig. 3. Framework for Web Data Minig

1.4 Social data mining

Social data mining is a new and challenging aspect of data mining. It is a fast-growing research area, in which connections among and interactions between individuals are analyzed to understand innovation, collective decision making, problem solving, and how the structure of organizations and social networks impacts these processes. Social data mining includes various tasks such as the discovery of communities, searching for multimedia data (images, video, etc) personalization, search methods for social activities (find friends), text mining for blogs or other forums. Social data mining finds several applications; for instance, in e-commerce (recommender systems), in multimedia searching (high volumes of digital photos, videos, audio recordings), in bibliometrics (publication patterns) and in homeland security (terrorist networks).

Social data mining systems enable people to share opinions and obtain a benefit from each other's experience. These systems do this by mining and redistributing information from computational records of social activity such as Usenet messages, system usage history, citations, and hyperlinks among others. Two general questions for evaluating such systems are: (1) is the extracted information valuable? , and (2) do interfaces based on extracted information improve user tasks performance?.

Social data mining approaches seek analogous situations in the computational world. Researchers look for situations where groups of people are producing computational records (such as documents, Usenet messages, or web sites and links) as part of their normal activity. Potentially useful information implicit in these records is identified, computational techniques to harvest and aggregate the information are invented, and visualization techniques to present the results are designed. Figure 4. Shows a traditional Data mining process.Thus, computation discovers and makes explicit the "paths through the woods" created by particular user communities. And, unlike ratings-based collaborative filtering systems (Hill & Terveen, 1996)., social data mining systems do not require users to engage in any new activity; rather, they seek to exploit user preference information implicit in records of existing activity. The "history-enriched digital objects" line of work (Resnick et al., 1994) was a seminal effort in this approach. It began from the observation that objects in the real world accumulate wear over the history of their use, and that this wear — such as the path through the woods or the dog-eared pages in a paperback book or the smudges on certain recipes in a cookbook — informs future usage. Edit Wear and Read Wear were terms used to describe computational analogies of these phenomena. Statistics such as time spent reading various parts of a document, counts of spreadsheet cell recalculations, and menu selections were captured. These statistics were then used to modify the appearance of documents and other interface objects in accordance with prior use. For example, scrollbars were annotated with horizontal lines of differing length and color to represent amount of editing (or reading) by various users.

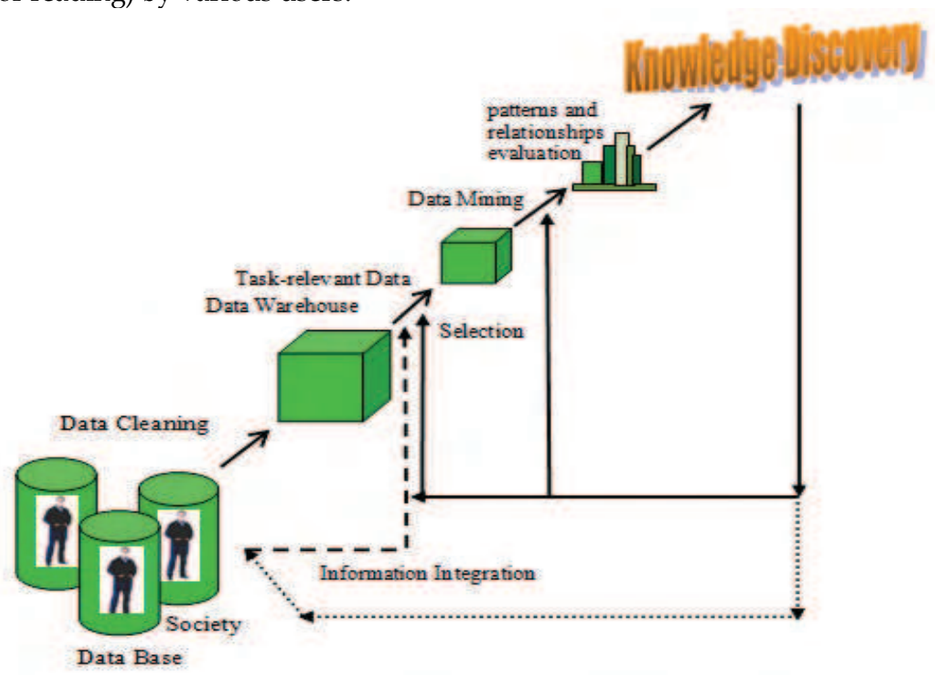

Fig. 4. A traditional Social Data Mining process (Ochoa, 2006).

The examples above mentioned are activities to which we are exposed and that without knowing we make use of the Data Mining, Due to this reason in the last years, Data Mining has had had great advances in artificial intelligence in order to offer a better support to user task (Ochoa, 2006).

2. Social networks

Web communities have risen rapidly in recent years with benefits for different types of users. For individuals, the Web community helps the users in finding friends of similar interests, providing timely help and allowing them to share interests with each other. For commercial advertisers, they can exploit the Web community to find out what the users are interested on, in order to focus their targets. It would be straightforward to discover the Web community if we had the detailed and up-to-date profiles of the relations among Web users. However, it is not easy to obtain and maintain the profiles manually. Therefore, the automatic approaches in mining users' relationship are badly needed.

Social network describes a group of social entities and the pattern of inter-relationships among them. What the relationship means varies, from those of social nature, such as values, visions, ideas, financial exchange, friendship, dislike, conflict, trade, kinship or friendship among people, to that of transactional nature, such as trading relationship between countries. Despite the variability in semantics, social networks share a common structure in which social entities, generically termed actors, are inter-linked through units of relationships between a pair of actors known as: tie, link, or pair. By considering as nodes and ties as edges, social network can be represented as a graph.

A social network is a social structure made of nodes (which are generally individuals or organizations) that are tied by one or more specific types of interdependency (See Fig. 5).

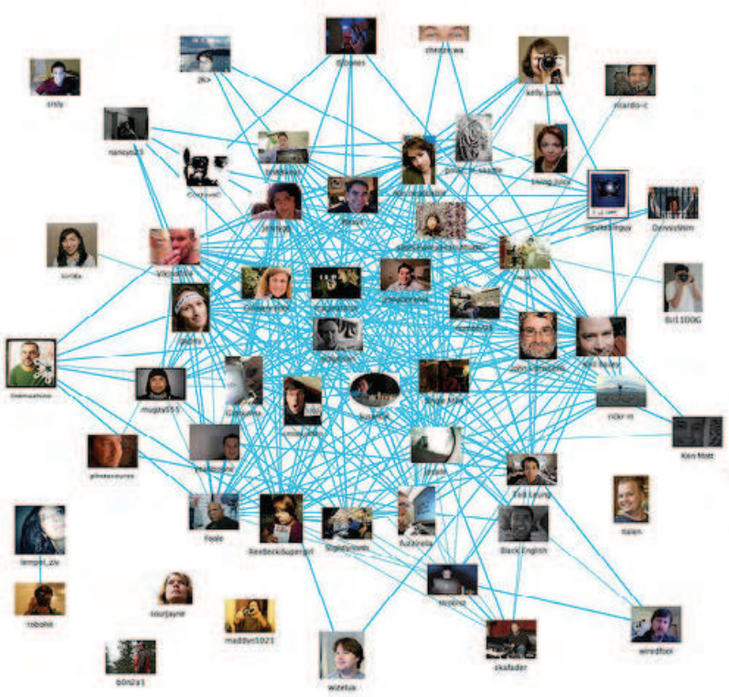

Fig. 5. Social Network Diagram.

2.1 Social networks analysis

Social Network Analysis (SNA) became a hot research topic after the seminal work by (Milgram, 1967). SNA is the study of mathematical models for relationships among entities such as people, organizations and groups in a social network. The relationships can be various. For example, they can be friendship, business relationship, and common interest relationship. A social network is often modelled by a graph, where the nodes represent the entities, and an edge between two nodes indicates that a direct relationship exists between the two entities. Some typical problems in SNA include discovering groups of individuals sharing the same properties (Schwartz & Wood, 1993) and evaluating the importance of individuals (Domingos & Richardson, 2001). Previously, the research in the field of SNA has emphasized binary interaction data, with direct and/or weighted edges (Lorrain & White, 1971) and focused almost exclusively on very small networks, typically, in the low tens of entities (Wasserman & Faust, 1994).

Moreover, only considering the connectivity properties of networks without leveraging the information of the entities limits the application of SNA.

Social network analysis has attracted much attention in recent years. Community mining is one of the major directions in social network analysis. Most of the existing methods on community mining assume that there is only one kind of relation in the network, and moreover, the mining results are independent of the users' needs or preferences. However, in reality, there exist multiple, heterogeneous social networks, each representing a particular kind of relationship, and each kind of relationship may play a distinct role in a particular task. Thus mining networks by assuming only one kind of relation may miss a lot of valuable hidden community's information and may not be adaptable to the diverse information needs from different users (Cai et al., 2005).

A social network can be analyzed for many useful insights. For instance, the important actors in the network, those with more connections, or the greatest influence, can be found. Alternatively, it may be the connection paths with actors that are of interest. Analysts may look for the shortest paths, or the most novel types of connections. Sometimes, the focus may even be on finding subgroups that are especially cohesive or interesting.

Knowledge of social networks is useful in various application areas. In law enforcement concerning organized crimes such as drugs and money laundering or terrorism, knowing how the perpetrators are connected, would assist the effort to disrupt a criminal act or to identify suspects. In commerce, viral marketing exploits the relationship between existing and potential customers to increase sales of products and services. Members of a social network may also take advantage of their connections to meet other members, for instance through web sites facilitating networking or dating among their users (Lauw et al., 2005).

3. Web radio

A research is detailed to acquire knowledge about how to develop a Web Radio using Social Data Mining and Cultural Algorithms (Reynolds, 1998), to a best functionality of it. Main thematic of the web radio is music to dance, that includes all the rates that consider equipment to dance, its directed to an ample segment of the society, whose only restriction is focused towards the different musical likes, any person who has desires to listen music and why not, to take advantage of it to dance, doesn't matter sex, age, civil state, nationality or many other factors, can access to our web counting on access to internet. Data Mining is very useful to any kind of projects, for that reason, we decided to use it inside, with this

web, users and developers can interact among them easily. We can help users using Data Mining in the creation of theirs lists of songstaking into account previous experiences with the same characteristics for new users, according to the classification that belongs to it according to the information that it provides in it is user profile. In order to obtain that the user stay on line into our Web, we have many rewards for them, agreement with their localization and the number of hours that they stay on our site.

3.1 Web radio introduction

Throughout history, the advance of the technology is in constant growth, one of the greatest discoveries has been the Internet that has facilitated the growth of other technologies as well as, thanks to this, we can practically be in contact with any people with the entire world, and ensure communications between the societies.

The radio has been another mass media between the people, also it has had changes through the time. The radio was one of first mass media with which the society had contact, by means of them emitted news, music and soap opera radio. The network has supposed a significant change in the way of transmission of this media, and has caused the birth of stations that they exclusively emit through them.

3.2 Problem outline

Why do you think that the radio has turned upside down so much with Internet? Because thanks to different services (World wide web, electronic mail, the news, internet, chat, among others) of internet, it is possible to undergo with other forms of information and expression that go beyond the wireless sound and to incorporate, therefore, new contents. In addition, also is feasible to generate new forms of consumption and relation that a listener can have with means (Hill & Terveen, 1996), (Ochoa et al., 2006).

In order to be able to implement the Data Mining it was decided to do a web radio using diverse tools that the new technology provides, as well as different software types to facilitate the work in its creation, in order to practice and to know more about the behavior of the human beings before this type of technologies. The design and development of the Web Radio called "Wave Radio" allowed increasing knowledge in different areas from science and technology. The creation of the Web Radio is a tool That was designed to investigate the user's behaviour of the same. Also to be able to integrate user groups (clusters) according to a stable classification that explains the tastes, characteristics that they share to each other.

In a traditional interactive application there is not factor during the design and development process (Brooks, 1994). If a web radio is considered as an interactive application on line, then it requires of human factors coming from Human Computer-Interaction (HCI) area (Nielsen & Loranger, 2006).

4. Security in web applications

With the rapid growth of interest in the Internet, network security has become a major concern to companies throughout the world. The fact that the information and tools needed to penetrate the security of corporate networks are widely available has increased that concern. Data mining has been loosely defined as the process of extracting information from large amounts of data. In the context of security, the information we are seeking is the

knowledge of whether a security breach has been experienced, and if the answer is yes, who is the perpetrator (Barbará & Jajodia 2002). Among well known criminals are:

A. **Hackers.** Hackers are criminals who try to break into your computer system and steal your personal information or cause other troubles.
B. **Online advertising impostors.** Online marketing techniques may be used to trick you or your family into doing something that may have a negative outcome.
C. **Online predators.** These are usually adults who are interested in grooming children online for their own sexual pleasure.
D. **Identity theft.** Criminals can steal your personal information and pretend they are you for financial benefit.

Other types of online crime also exist where people can obtain a financial advantage illegally. Because of this increased focus on network security, network administrators often spend more effort protecting their networks than on actual network setup and administration. Tools that probe for system vulnerabilities, such as the Security Administrator Tool for Analyzing Networks (SATAN), and some of the newly available scanning and intrusion detection packages and appliances, assist in these efforts, but these tools only point out areas of weakness and may not provide a means to protect networks from all possible attacks. Thus, as a network administrator, you must constantly try to keep abreast of the large number of security issues confronting you in today's world. Data Mining in Web Security concentrates heavily in the area of intrusion detection.

Private information can reside in two states on a network. It can reside on physical storage media, such as a hard drive or memory, or it can reside in transit across the physical wired or wireless network in the form of packets. These two information states present multiple opportunities for attacks from users on your information, as well as those users on the Internet. We are primarily concerned with the second state, which involves network security issues. The following are five common methods of attack that present opportunities to compromise the information on your network:

- Network packet sniffers
- IP spoofing
- Password attacks
- Distribution of sensitive internal information to external sources
- Man-in-the-middle attacks

When protecting your information from these attacks, your concern is to prevent the theft, destruction, corruption, and introduction of information that can cause irreparable damage to sensitive and confidential data. According (Barbará & Jajodia 2002) the use of data mining is based on two important issues. First, the volume of data dealing with both network and host activity is so large that it makes it an ideal candidate for using data mining techniques. Second, intrusion detection is an extremely critical activity.

What is Network Intrusion Detection?

Intrusion detection starts with instrumentation of a computer network for data collection. Pattern-based software 'sensors' monitor the network traffic and raise 'alarms' when the traffic matches a saved pattern. Security analysts decide whether these alarms indicate an event serious enough to warrant a response. A response might be to shut down a part of the network, to phone the internet service provider associated with suspicious traffic, or to simply make note of unusual traffic for future reference. Intrusion detection systems are software and/or hardware components that monitor computer systems and analyze events occurring in them for signs of intrusions (Kumar et al., 2005).

Data mining based intrusion detection techniques generally fall into one of two categories; misuse detection and anomaly detection. In misuse detection, each instance in a data set is labelled as 'normal' or 'intrusion' and a learning algorithm is trained over the labelled data.
A key advantage of misuse detection techniques is their high degree of accuracy in detecting known attacks and their variations. Their obvious drawback is the inability to detect attacks whose instances have not yet been observed (Dokas et al., 2002). Anomaly detection, on the other hand, builds models of normal behaviour, and automatically detects any deviation from it, flagging the latter as suspect. Anomaly detection techniques thus identify new types of intrusions as deviations from normal usage (Javitz & Valdes 1993)

4.1 Derived data for intrusion detection

A single connection between an outside machine and a single port on a machine inside your network is not malicious—unless it is part of a series of connections that attempted to map all the active ports on that machine. For this reason you will want to add additional fields containing values from the base. Example, you could distinguish traffic originating from outside your network from traffic originating inside your network. Another type of derived data, called an aggregation, is a summary count of traffic matching some particular pattern. Example, we might want to know, for a particular source IP X, and a particular IP Y, how many unique destinations IP were contacted in a specific time window Z. A high value of this measure could give an indication of IP mapping, which is a pre-attack reconnaissance of the network. Aggregations are generally more expensive to compute than other kinds of derived data that are based upon only a single record. A third type of derived data is a flag indicating whether a particular alarm satisfies a heuristic rule. Because data mining methods handle many attributes well, and because we don't know for sure which one will be useful, our approach is to compute a large number of attributes (over one hundred) and store them in the database with the base alarm fields (Bloedorn et al., 2001).
Due to widespread diversity and complexity of computer infrastructures, it is difficult to provide a completely secure computer system. There are numerous security and intrusion detection systems that address different aspects of computer security. Below we present a common architecture of intrusion detection systems and its basic characteristics.

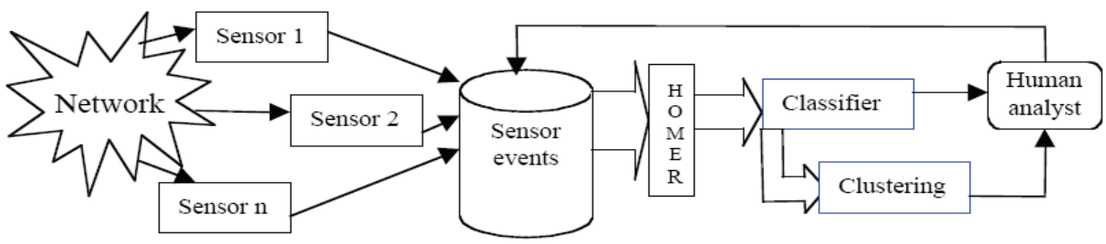

Fig. 6. How sensors feed into overall intrusion detection system (Bloedorn et al., 2001)

The figure above shows a proposed configuration to perform intrusion detection. First, network traffic is analyzed by a variety of available sensors. This sensor data is pulled periodically to a central server for conditioning and input to a relational database.
Before security specialists can start providing input to the data mining effort, this traffic must be filtered. It is a straightforward task to create a filter that can find these patterns within a data table of traffic. At figure 6, this preliminary filter is called HOMER (Heuristic

for Obvious Mapping Episode Recognition). The heuristic operates on aggregations by source IP, destination port, and protocol and then check to see if a certain threshold of destination IPs were hit within a time window. If the threshold is crossed, an incident is generated and logged to the database.
Even though the bulk traffic due to the mapping activity is not shown to the analyst, the source host itself is placed on the radar screen of our system. Please note that some normal activity (e.g., name servers, proxies) within an organization's intranet can match the profile of an IP mapping. HOMER handles this situation by means of an exclusion list of source IPs.
HOMER filters events from the sensor data before they are passed on to the classifier and clustering analyses. Data mining tools filter false alarms and identify anomalous behaviour in the large amounts of remaining data. A web server is available as a front end to the database if needed, and analysts can launch a number of predefined queries as well as free form SQL queries from this interface. The goal of this operational model is to have all alarms reviewed by human analysts.
Catching new attacks can not depend on the current set of classification rules. Since classification assumes that incoming data will match that seen in the past, classification may be an inappropriate approach to finding new attacks. K-means clustering is used to find natural groupings of similar alarm records. Records that are far from any of these clusters indicate unusual activity that may be part of a new attack. Finally, we discussed a hot subject on web security: preserving privacy data mining.

4.2 Preserving Privacy Data Mining (PPDM)

What is privacy?
In (Shoeman, 1984) was defined privacy as "the right to determine what (personal) information is communicated to others" or "the control an individual has over information about himself or herself." More recently, Garfinkel (Garfinkel, 2001) stated that "privacy is about self-possession, autonomy, and integrity."
Another view is corporate privacy – the release of information about a collection of data rather than an individual data item. For example: "I may not be concerned about someone knowing my birth date, mother's maiden name, or social security number; but knowing all of them enables identity theft" (Clifton et al., 2004).

4.3 Knowledge discovery and privacy

When people talk of privacy, they say "keep information about me from being available to others". However, their real concern is that their information not be misused. The fear is that once information is released, it will be impossible to prevent misuse. Utilizing this distinction – ensuring that a data mining project won't enable misuse of personal information – opens opportunities that "complete privacy" would prevent (Clifton et al. 2004). The key finding is that knowledge discovery can open new threats to informational privacy and information security if not done or used properly (Oliveria & Zaiane, 2004).
Privacy is viewed as a social and cultural concept. However, with the ubiquity of computers and the emergence of the Web, privacy has also become a digital problem (Rezgur et al. 2003). With the Web and the emergence of data mining, privacy concerns have posed technical challenges different from those that occurred before the information era.
In data mining the definition of privacy preservation is still an unclear topic. A notable exception is the work presented in (Clifton et al. 2002), in which PPDM is defined as

"getting valid data mining results without learning the underlying data values." According to (Oliveria & Zaiane, 2004) PPDM encompasses the dual goal of meeting privacy requirements and providing valid data mining results. This definition emphasizes the dilemma of balancing privacy preservation and knowledge disclosure.

5. Internet frauds

Fraud is the crime or offense of deliberately deceiving another in order to damage them usually, to obtain property or services unjustly. Fraud can be accomplished through the aid of forged objects. In the criminal law of common law jurisdictions it may be called "theft by deception", "larceny by trick," "larceny by fraud and deception" or something similar.

Sentinel Top Complaint Categories January 1 – December 31, 2007			
Rank	Top Categories	Complaints	Percentage
1	Identity Theft	258,427	32%
2	Shop-at-Home/Catalog Sales	62,811	8%
3	Internet Services	42,266	5%
4	Foreign Money Offers	32,868	4%
5	Prizes/Sweepstakes and Lotteries	32,162	4%
6	Computer Equipment and Software	27,036	3%
7	Internet Auctions	24,376	3%
8	Health Care	16,097	2%
9	Travel, Vacations and Timeshare	14,903	2%
10	Advance-Fee Loans and Credit Protection/Repair	14,342	2%
11	Investments	13,705	2%
12	Magazines and Buyers Clubs	12,970	2%
13	Business Opps and Work-at-Home Plans	11,362	1%
14	Real Estate (Not Timeshare)	9,475	1%
15	Office Supplies and Services	9,211	1%
16	Telephone Services	8,155	1%
17	Employ Agencies/Job Counsel/Overseas Work	5,932	1%
18	Debt Management/Credit Counseling	3,442	<1%
19	Multi-Level Mktg/Pyramids/Chain Letters	3,092	<1%
20	Charitable Solicitations	1,843	<1%

Table 1. Internet Frauds made in January-December 2007

Internet fraud generally refers to any type of fraud scheme that uses one or more online services - such as chat rooms, e-mail, message boards, or Web sites - to present fraudulent solicitations to prospective victims, to conduct fraudulent transactions, or to transmit the proceeds of fraud to financial institutions or to others connected with the scheme. Unfortunately, people who engage in fraud often operate in "Internet time" as well. They seek to take advantage of the Internet's unique capabilities -- for example, by sending e-mail messages worldwide in seconds, or posting Web site information that is readily accessible

from anywhere in the world - to carry out various types of fraudulent schemes more quickly than was possible with many fraud schemes in the past.

The types of Internet Fraud, in general, the same types of fraud schemes that have victimized consumers and investors for many years before the creation of the Internet are now appearing online (sometimes with particular refinements that are unique to Internet technology). With the explosive growth of the Internet, and e-commerce in particular, online criminals try to present fraudulent schemes in ways that look, as much as possible, like the goods and services that the vast majority of legitimate e-commerce merchants offer. In the process, they not only cause harm to consumers and investors, but also undermine consumer confidence in legitimate e-commerce and the Internet. There are some of the major types of Internet fraud that law enforcement and regulatory authorities and consumer organizations are seeing in the USA:

- Auction and Retail Schemes Online. This type of fraudulent schemes appearing on online auction sites are the most frequently reported form of Internet fraud. These schemes, and similar schemes for online retail goods, typically offer high-value items - ranging from items that are likely to attract many consumers. These schemes induce their victims to send money for the promised items, but then deliver nothing or only an item far less valuable than what was promised (e.g., counterfeit or altered goods).
- Business Opportunity/"Work-at-Home" Schemes Online. Fraudulent schemes often use the Internet to advertise purported business opportunities that will allow individuals to earn thousands of dollars a month in "work-at-home" ventures. These schemes typically require the individuals to pay a guaranteed amount of money, but fail to deliver the materials or information that would be needed to make the work-at-home opportunity a potentially viable business.
- Identity Theft and Fraud. Some Internet fraud schemes also involve identity theft - the wrongful obtaining and using of someone else's personal data in some way that involves fraud or deception, typically for economic gain.
- Investment Schemes Online
    - Market Manipulation Schemes. Enforcement actions by the Securities and Exchange Commission and criminal prosecutions indicate that criminals are using two basic methods for trying to manipulate securities markets for their personal profit. First, in so-called "pump-and-dump" schemes, they typically disseminate false and fraudulent information in an effort to cause dramatic price increases in thinly traded stocks or stocks of shell companies (the "pump"), then immediately sell off their holdings of those stocks (the "dump") to realize substantial profits before the stock price falls back to its usual low level. Second, in short-selling or "scalping" schemes, the scheme takes a similar approach, by disseminating false or fraudulent information in an effort to cause price decreases in a particular company's stock.

There are other Schemes of Internet Fraud besides before mentioned (http://www.usdoj.gov/criminal/fraud/internet/). The fraud detection is becoming increasingly important in revealing and limiting revenue loss due to fraud. Fraudsters aim to use services without paying or illicitly benefit from the service in other ways, causing service providers financial damage. To reduce losses due to fraud, one can deploy a fraud detection system. However, without tuning and through testing, the detection system may cost more in terms of human investigation of all the false alarms than the gain from

reduction of fraud. Test data suitable for evaluating detection schemes, mechanisms and systems are essential to meet these requirements. The data must be representative of normal and attack behaviour in the target system since detection systems can, and should, be very sensitive to variations in input data.

6. Diverse applications and domains where analysis through data mining

In this section we show some application related to the topics before mentioned.

6.1 Social networks application

Orkut is a system of social networks used in Brazil by 13 million users, many of them, create more of a profile, and generate different relationships from their different profiles, this takes to think that they develop Bipolar Syndrome, to be able to establish communications with people of different life styles, and when they doing to believe other users that they are different people (Zolezzi-Hatsukimi, 2007).

Some problems in the social network of Orkut exist that dislike much to their users. The loss of privacy, the lack of materialization of the relations established through the network. The false profiles are created for: to make a joke, to harass other users, or to see who visualizes its profile. As the profile is false, the friends of this profile are also generally false, making difficult the tracking of the original author. The users can make denunciations against those false profiles, but without clear the profile of Orkut. Anyway, the original author can create a new profile. Often a user does not wish to exhibit his photo in Orkut and put a photo of a celebrity. That is more common and is accepted by the community for those who wishes to remain anonymous. In this case, it uses his real name and places photos for the album, but with a photo of any profile.

The most serious event in Orkut is the creation of communities with a concept of racism, xenophobia and mistreat against the animals and making vindication to the consumption and sale of drugs and pedofilia. Unfortunately, the users who denounce these facts to be eliminated do not reach their objective: the criminals create these data again, deceiving to the server of Orkut. Nevertheless, due to pressures on the part of the Brazilian government and of the American press, new actions on the matter of this were announced on the part of the server, in a definitively effective action.

Many critics are made against Orkut. The main one is the libertinism of that place, can be spoken on racism, murders, without nothing happens, if the person knows to hide it. The moderation in Orkut is of ear, nevertheless is not sufficient. The police tries to find guilty of certain crimes, since some of them agree all through Orkut. To some users its own profile is kidnapped, rob password of email, MSN or banking accounts.

Using the tool of Data Mining denominated WEKA, it was come to develop a de-nominated Model "Ahankara" of prediction of profiles in users of Orkut, which al-lows to understand the motivations of this type of profile and to determine if it has generated Syndrome Bipolar, to see figure 3 (Ponce et al., 2007).

The model obtained Ahankara once used WEKA to look for the relations that us could be of utility to process the data. see Figure 7.

Some of the relations that obtained when observing the data with aid of the WEKA are the following ones.

- The region has to do, with the number of fans.

- The number of communities is based on the number of people with interests in common, some factors can take part like sex and civil state.
- Regions exist in which there is people in all the communities, regions exist in which the people of that region this in a single community.
- The people who participate in many communities, the majority have less fans, so that in theory they spend long time in entering all the communities to which they have themselves you incorporate.
- The Region if it influences in the number of communities which the people enter, due to the activities who are used to doing has certain interests in common in each region.

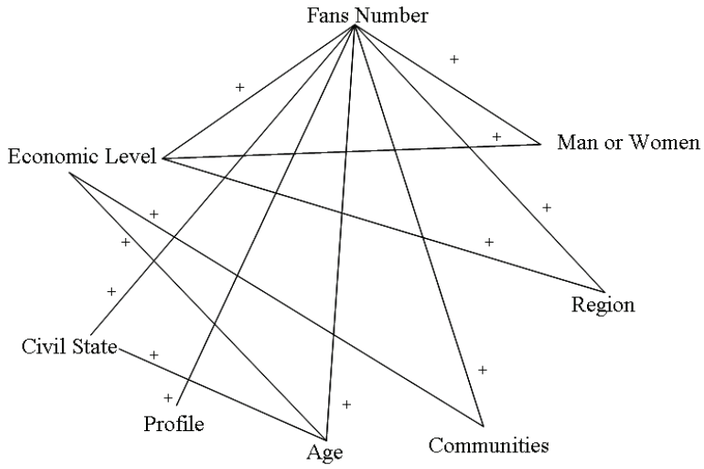

Fig. 7. Ahankara Model, This is the propose model (Ponce et al., 2007).

Through Mining Data it is possible to find relations between the data, often these can be hidden and others are evident, another type of analyses that can be realised are especially the groups of people by some type of characteristics, this can be realised through diverse techniques of clusters or heuristic technical using as is the used in (Ponce et al., 2006) where the clique maximum problem is solved, which can used to find the clusters with major number to relationship in bases of a certain characteristic.

6.2 Web radio application

6.2.1 Data mining and cultural algorithms for develop the intelligent web radio

We purpose to develop web radio applications taken into account in particular social acceptability factor using Social data mining and cultural algorithms. By so doing, we purpose a conceptual model which include the operation of the web radio, everything begins at the moment in which the user creates his profile, after for creating his data are incorporate in a data base in which all the profiles are stored, to be able to make the analysis with them, so that if at some time a user with different preferences arrives and that he does not share with the other users already registered it stores a new case in the case library so that in the future it can be reused and help the users to create their lists of songs on the basis of its profile. And this process is repeated whenever a new user enters his/her information; this is shown in figure 3.

When people listen to music they do not like, their initial reaction is to fast forward, followed by changing moods if they do not hear acceptable music within a reasonable

number of fast forwards. We believe that people appreciate having these two options. This makes our selection mechanism different from radio. Using in the web radio with the cultural algorithms, it could improve the performance because it is possible to motivate a society with different preferences and analyze the user preferences. With the cultural algorithms is possible analyze what kind of music is someone listening, our Web Radio can deduce the songs, the singers and the genders the person prefers, and by using this information recommend additional songs (see Figure 8).

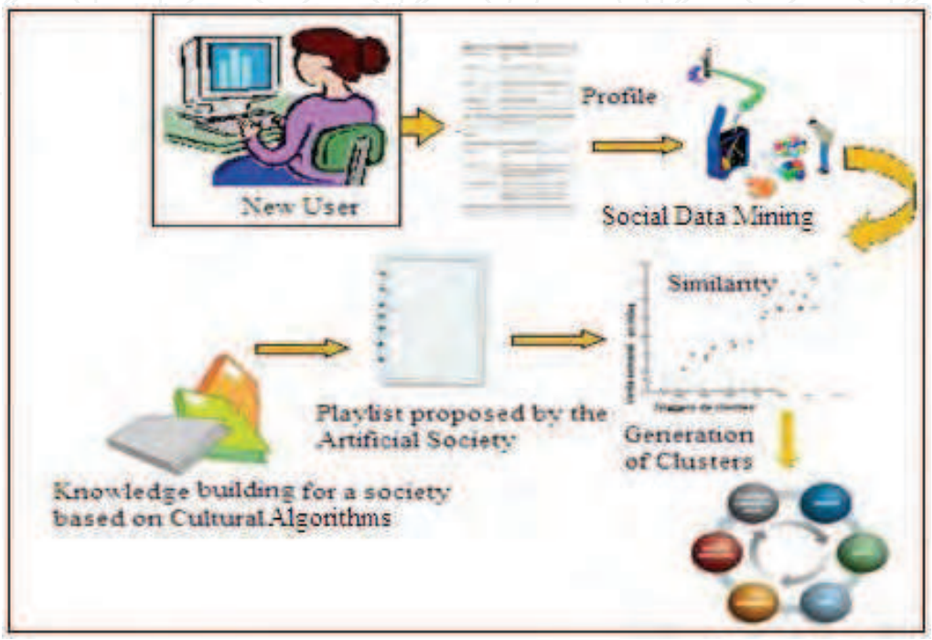

Fig. 8. Design and development of a Web radio based on Social Data Mining and cultural algorithms.

We made a system that allows users to view individual and group historical listening lists and define his new lists. These systems learn of the user preferences, after songs are selected to be played on a shared physical environment, based on the preferences of the whole people present, similar to Tibetan Avenue (http://tibetanavenue.com/).

According to users profile, They are classified of a way in which those that occupies that group feel identified with other users who perhaps are not of the same country but who share other characteristics such as; age, musical sort, zodiacal sign. When grouping of this form to the users groups with an almost equal personality. One says almost because they do not share all the characteristics, but in his majority they have the same pleasures. In the social network there is one or two people whom a greater number of features with other users and there shares it is where it is the base of the social network.

An extra on which it tells to the Web Radio is that it has prizes of reward for the users who spend major time in tuning, this according to the region where is being in tune and also vary with the participation that has within the Web Radio. For it every month related to the musical sorts will be realized one trivia on which it tells to the Web Radio. If the user has a major number of options to answer correctly, he/her will be made creditor have access to unload a video of steps to learn to dance, the users only can answer one trivia per week.

6.2.2 A thematic web radio application

In general a music delivery system is classified in two broad categories, content purchasing and audio broadcasting. In the case of content purchasing, the consumer pays for specific music (e.g. CDs, cassette tapes, LPs) to build up a collection of personal favorites over which he/she has complete control. Broadcast audio (e.g. radio, TV, Internet radio) provides more content, but at the cost of limited consumer control (consumers choose radio stations, not music). Personalized audio music attempts to bridge the gap between the two. Our hypothesis is that by matching information about the users (listener profiles) with the knowledge building for a society based on cultural algorithms (content metadata), it should be possible to automatically generate more pertinent playlists for individual listeners. In this paper we test this hypothesis.

Thematic web radio is a web application developed using the programming language PHP by means of the Macromedia family was developed the graphical interface so that it is more efficient, pleasant and easy the handling of the same for the user (Field et al., 2001) .

Within the diverse functions that realize the system is had as primary target the handling, administration, search and analysis of users, profiles and tastes, among others features, similar at previous research, which was developed a prototype (Ochoa et al., 2007).

One of the functions that support in the good operation of the web radio is, by means of the use of a profile, the user can provide data of personal interests to the creators of the Web such as: sex, age, date of birth, e-mail, areas of interest, among others, and so it is possible to be grouped to the users according to the characteristics that share to generate clusters.

The registered users can have access to the creation of a play list, at the time of which the user creates his/her list, of the songs on which he/she tells to the data base of the Web Radio.

The list is stored in a data base and this can be used by other users, since, to give pursuit to the social network, we must have like minimum ten cases, so that we pruned to follow with the implementation of the Social Data Mining (see Figure 9).

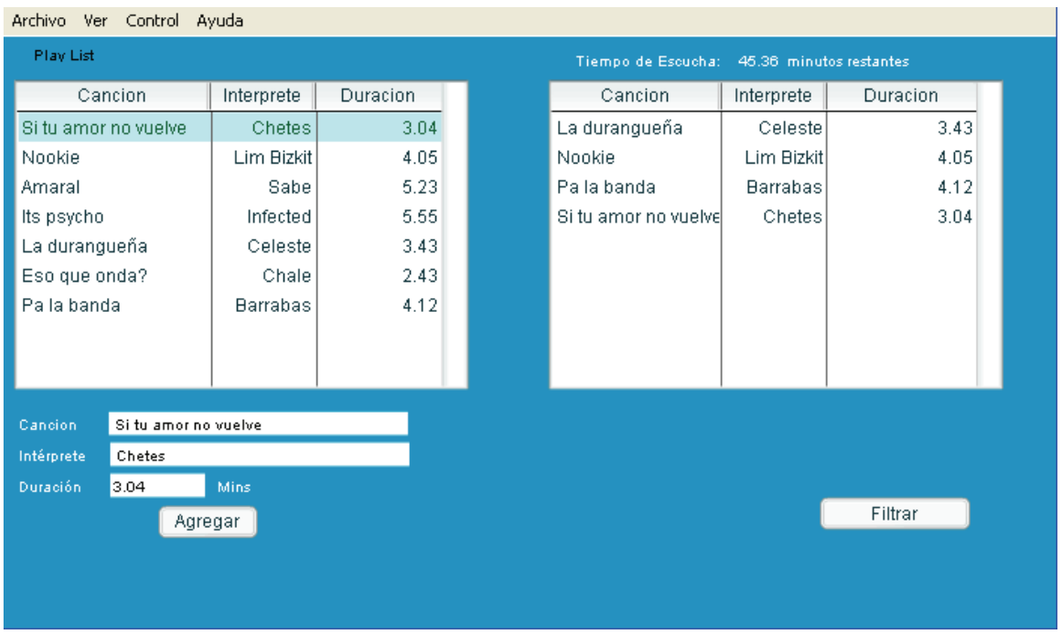

Fig. 9. Play list user interface.

Another function of the Web Radio is a finder that allows to locate songs, the searches are based on hierarchies such as title, interprets, among others. By means of the finder one facilitates the use and the permanence of the user.
Within the operation of the web radio is a function that allows that the users has power by deciding the ranking of the songs according to its preference, the scale that is used for the measurement is the "Lickert Scale" that is defined like a series of items or phrases that carefully have been selected, so that they constitute a valid criterion, trustworthy and precise to measure of some form the social phenomena.
The functions before mentioned allow having a better control and thus we can give a pursuit to the social network to which our Web Radio is associated. The profiles are stored and with them we can compute the range between the people using theirs profiles (to look for the similarities) and in this way of specifying the limits of clusters (Users with similar preferences). We use the similarity function used in Social Data Mining to organize the people in different clusters:

$$\frac{\sum_{i=1}^{n} w_i \times sim\left(f_i^I, f_i^R\right)}{\sum_{i=1}^{n} w_i} \tag{1}$$

In where:
w_i is the weight of importance about an attribute.
sim is the function of similarity.
f_i^I, f_i^R are the values of attribute i in the entrance profile (I) and the recover profile (R).

6.2.3 Experiments related with the artificial social net that support this web radio

We have worked with two scenarios, where web users use our thematic web radio. In the first scenario, we compared the performance of 27 communities of 50 agents, and on the other hand 27 communities of 500 agents each one. The optimal number of songs in the playlist was 17. One of the most interesting characteristics observed in this experiment is the diversity of cultural patterns established for each community. For the solution with the same number of songs the provided result for the "beliefspace" is entirely different. The structured scenarios associated to the agents cannot be reproduced in general due they belong to a given instant in the time and space.
They represent a unique, precise and innovative form of adaptive behavior which solves a computational problem followed by a complex change of relationships. The generated configurations can be metaphorically related to the knowledge of the community behavior regarding to an optimization problem (to decide the music related with the playlist), or a tradition with which to emerge from the experience and with which to begin a dynamics of the process [1]. Comparing the 50 agents of the first community regarding the 500 agents community, this last obtained a better performance in terms of the average songs (17.05 versus 18.30), as well as a smaller standard deviation (1.96 versus 2.64). They also had a greater average number of changes in the paradigm (5.85 versus 4.25), which indicates that even the "less negotiating" generations. In the second experiment, we consider the same scenario for the experiment one, except that after having obtained a solution from a community of 50 agents. In this experiment, it was surprising to see initially how the community of 500 agents uses the solution offered by the 50 agents, whenever these

solutions were close the optimal grade, instead of finding entirely complete new solutions. This can be compared metaphorically with the concept of culture.

6.3 Using biometry and data mining in online assessments to detect who is there?

This application is a clear example of the use of data mining on security, by means of the system described is possible to avoid impersonation as well as academic frauds in online assessments. Unless photo IDs are checked and all course work occurs inside of a monitored classroom, faculty really does not know for sure whether the student is who they say they are in the classroom or online (MSU, 2006). On online assessments in which we are not sure who is taking the test; students will be under pressure, some students perform unfairly poorly under pressure and this is a good incentive to cheat (Rove, 2004). We have a wide spectrum of documented techniques to commit cheat on online assessments: modify a grade in the database (DB), to steal answers for questions, to copy from another student or cheat sheets, impostor or substitute remote students, to search for answers on the Internet or in blogs or purchase the list of answer for an specific exam, on the messenger or cellular phone, in single words to "commit cheat" to obtain a "better grade" in an online assessment. Biometrics is becoming a powerful tool to improve security on transactions and reduce frauds (ITEDU, 2006).

An advanced security measure can be implemented by means of biometric technologies; much of the hot discussion about biometrics has come about due to the level of research and interest shown in large scale implementations of the technology by the US and UK Governments and the European Union (Clarke & Furnell 2005). They may provide added robustness in access control to high security facilities within higher education. As the unit price for biometric devices continues to fall is possible to employ these to replace the current systems used for workstation and network access (Wasniowski, 2005). These devices are likely to become a standard computer peripheral, built into future workstations.

6.3.1 The problem

The main problem on online assessments is to know who's there (Wisher et al. 2005). In this section, we propose the use of biometrics and data mining, particularly the use fingerprint recognition and web cam monitoring on real time to verify student's identity during online assessment; we propose also the analysis of the student's behavioural patterns by means of data mining to deal with the well-known problem of: who is taking the exam? The contribution of this paper is the use of hybrid technologies in online assessments as a new approach for remote identification of students on real time.

6.3.2 Performance schema (3-tier client-server system)

We separated the application in three main modules: the first one is on charge of the conduction the online assessment, the second one on charge of the fingerprint recognition and web cam monitoring on real time, the third on charge of data mining analysis on students behavioural patterns. Server must be in listening mode waiting for Clients that requires a service. In order to use fingerprint recognition, the first step is to enrol students – top, right side in Figure 10-, the students fingerprint is saved and indexed in the Features Database, we highly recommend to separate this from the Assessment System Database, using even separated servers, to improve system overall performance. In the features

database is assigned the Student Personnel ID that is used to link the students' personnel information with the fingerprint image.

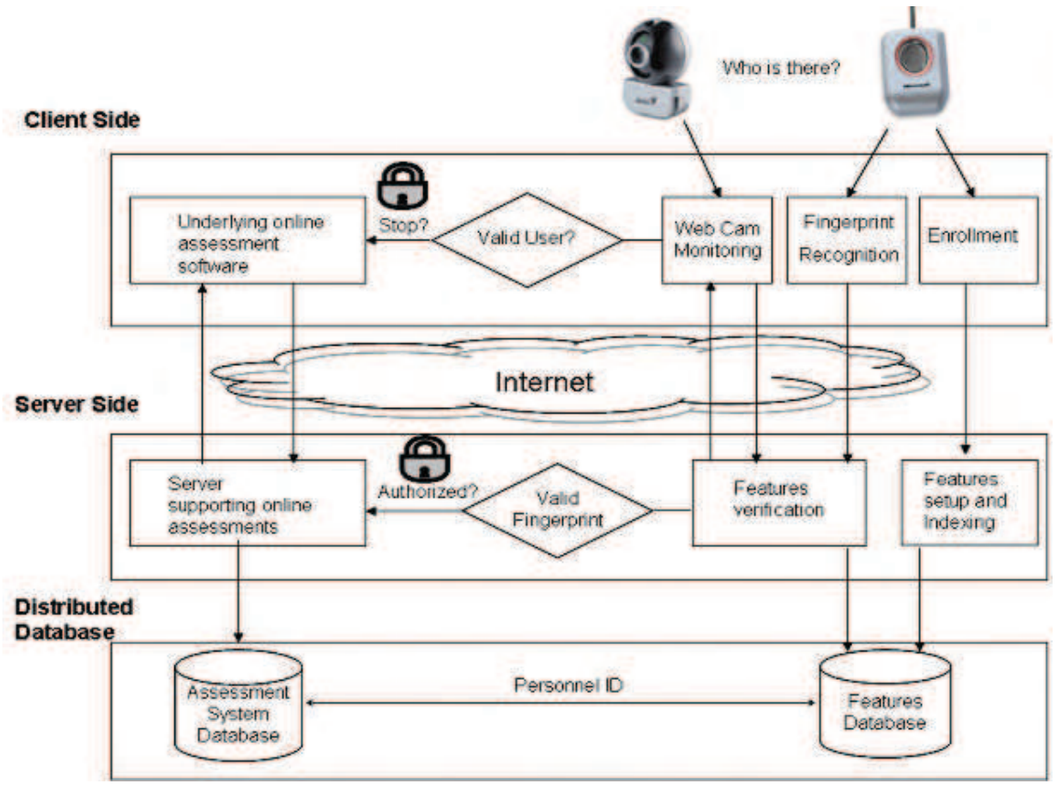

Fig. 10.- Student's biometric recognition on real time & data mining analysis

6.3.3 Implementation

- Hardware
  - Client System Requirements (minimal). Pentium class (i386) processor (200 MHz or above) with 128Mb or higher, 100Mb disk space.
  - Fingerprint mouse. 250 DPI (Digits per Inch) or higher; 500 DPI is recommended.
  - Web cam. Genius VideoCaM GE111 or VideoCam GF112.
  - Broad-band Internet. Minimum 128 Kbps, recommended 256 Kbps.
- Software
  - Biometrics SDK. Griaule GrFinger SDK 4.2 allows you to integrate biometrics in a wide variety of applications. Provides Support for dozens of programming languages –including java- and integration with several Database Management Systems. Besides, provides multiple fingerprint reader support, and even after application development or deployment, makes you able to change the fingerprint reader you're using, without modifying your code.
  - Fingerprint template size: 900 bytes average.
  - RapidMiner (Version 4.1). To carry out data mining process.
  - Programming language. Java due the online assessment software tool was developed using this technology, and JMF (Java Media Framework) to allow transmission of video and/or photographs over the Internet.

- Web Server. Apache 2.2.
- Database Management System. MySQL.
- Operating System: Windows 98, Windows ME, Windows NT. Windows 2000, Windows XP, Windows 2003 or Windows Vista.

The fingerprint is verified in the Features Database, and if it is recognized as a valid, then the Server authorizes access to the online assessment application, else an error message is sent to the Client to try again. In other hand, if the student's fingerprint is valid, the user is authenticated into system (see Figure 11), the evaluation process starts and web cam transmission is initialized at Client Side to conduct real time monitoring by means of multitasking, students' activities (keystroke dynamics, navigation and performance patterns) are stored on features database and log files.

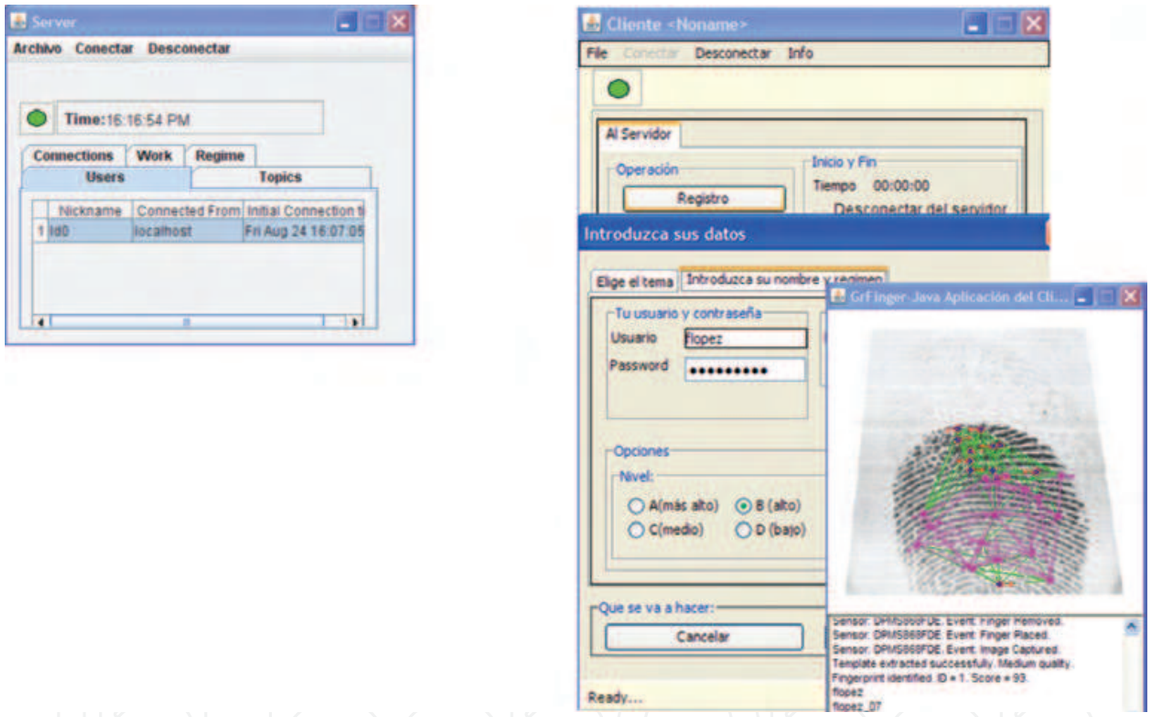

Fig. 11. The Client-Server Application supporting fingerprint recognition to authenticate students in online assessments

If someone else tries to get the control of the computer during the online assessment, the evaluation process is finished prematurely, and the results are sent to server side to be processed as they are. To the contrary, the evaluation process is finished successfully, the assessment is processed at Server Side, and the final results of evaluation and security status are shown at Client Side (Hernández et al. 2008). The stage of data mining is executed asynchronously to verify students' behaviour patterns: keystroke dynamics (Gutierréz et al. 2002), navigation patterns (Xing & Shen 2004) and performance patterns (Hernández et. al 2006). After repeatedly usage of the system, these combined patterns can be used to verify students' identity and even to substitute the usage of webcams.

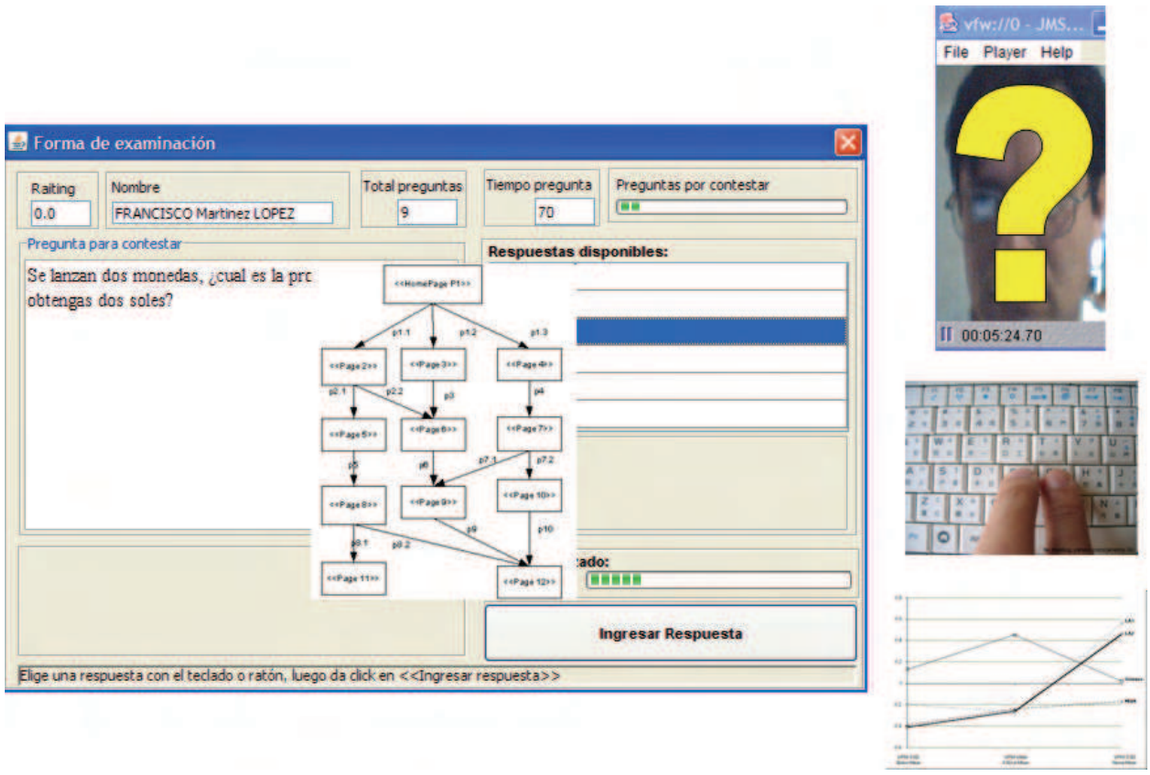

Fig. 12.- Assessment process: recording of keystroke, navigation and performance patterns.

6.3.4 Preliminary results

We develop an experiment to try the above mentioned technology. We selected a random sample of students (n=54) from the José María Morelos y Pavón High School, located in Temixco, Morelos, México. On this test with obtain a False Acceptance Rate (FAR) of 99.99% and a False Reject Rate (FRR) of 97.09%, only one student could not be recognized despite several trials, although we try enrolled her trying different fingers of her left hand, simply we could not, she has tiny long fingers and the enrolment results were always the same. Her fingerprint template can not be understood by the system due is confuse, her fingerprints templates seems like stains. Something related is registered in literature, Asiatic persons has similar problems to be identified by fingerprint readers (Michigan Org, 2007). We faced this problem by providing this student an user and a strong password to allow permission to the system.

In general, students perceived our system as faster, easy to use and secure; fingerprint recognition plays an important role in this last point. However 13% dislike web cam monitoring. When we asked them directly if they dislike being monitoring, 33% answered bothers this fact. They felt under pressure, get nervous and dislike being monitored or watched.

A 20% noticed a way to commit cheat using a system like ours, the ways are: turn the camera to some else, use a photo, use a cheating list, and just one person thinks to dirt the fingerprint reader. We made in-depth analysis and discover that students with poor performance (low grades) are willing to commit cheat. Finally, 78 % of the students would like the system being implemented at their high school.

Data mining process shows promising results to identify remote students by using keystroke dynamics; meanwhile navigation and performance patterns are useful to identify suspicious behaviour (i.e. unexplained grades, unidentified remote IPs, and completely different navigation patterns –regarding previous- when solving an exam).
We consider that the online assessment system with biometric recognition was very well accepted, but must be adapted to be more user friendly and the process to enrol users must be improved too.

7. Conclusion and the future research

The quickly grown of the Web has done that it is a great information source in many areas, which can be used to obtain important data in different areas like social, psychological, marketing, among others. As one saw from point psychological are possible to be studied some behaviors and found certain landlords with the help of Data Mining, also it is possible to determine future behaviors on the basis of certain antecedents as one is in the application on the basis of social networks information like Orkut. On the other hand it can predict certain preferences by some product or service; this could be observed through the Web Radio application. The Web Data Mining can help us to understand more some things and provide a base for the decision make. In the Web one can find a great amount of information sources, the unique thing that there is to think is those that are wanted to obtain. In this case our applications analyzed were Social Networks, Web radio, Security and Internet frauds. In this work to show a conceptual model to develop systematically web radio applications taking into account social acceptability factor using the social data mining and cultural algorithms. With the matching information about the users (listener profiles) with the knowledge building for a society based on cultural algorithms (content metadata), it could be possible to automatically generate more pertinent playlists for individual listeners. Then, it is feasible that a web radio purpose could interpret the human behaviour under certain situations to which it is exposed, this by means of the behaviour that will have the users when interacting with the Web Radio. This is only one part of everything what it is possible to be done with the aid of the Social Data Mining. With the creation of the web radio one hopes that one undertakes new projects that are developed with the aid of the Social Data Mining and cultural algorithms.
Nevertheless there are a lot of research work that will be doing with Data Mining for Web applications and some future works that can be: In the area of social networks it is the search of criminal networks in the diverse social networks like Orkut, MySpace, Badoo, Hi5, etc. these criminal networks are possible to be dedicated to the kidnapping, traffic of bodies, drug traffic, among others activities. Another application is to make an analysis of these networks to obtain information that can be used to offer products and services in specific sectors, an example of this is the Web radio, but it is possible to be analyzed another types of services with the security that these are going well to be received by a certain Web user group.
In the Internet area fraud we want to improve human-computer interface and assessment methodology by including student's comments and users feedback. We want to test the tool with different groups at different high schools and Universities. Regarding biometric recognition, we want to improve facial recognition, due at this point of our research we can

detect student's presence or absence only, and our intention is comparing face patterns automatically by means of photo Ids stored at our features databases. We want to test the newest fingerprint scanners included in mouses, keyboards and in some laptops and try to incorporate them to work within our system.

8. References

Bloedorn, E.; Christiansen, D.; Hill, W.; Skorupka, C.; Talbot, L. & Tivel J. (2001). Mining for Network Intrusion Detection: How to Get Started, *MITRE Technical Report*, August 2001

Brooks, P. (1994). Adding value to usability testing, In: *Usability Inspection Methods*, Nielsen, Jakob y Mack, Robert, 1, pp. 255-271, Published by John Wiley & Sons, New York, NY

Cai, D.; Shao, Z.; He, X.; Yan, X. & Han J. (2005). Mining Hidden Community in Heterogeneous Social Networks, *Proceedings of LinkKDD'05*, Chicago, USA, August 2005, ACM

Clarke, N. & Furnell, S. (2005). Biometrics: "The promise versus the practice", *In Computer Fraud & Security*, Volume 2005, pp. 12-16, September 2005

Clifton, C.; Kantarcioglu, M. & Vaidya J. (2002). Defining Privacy For Data Mining, *Proceedings of the National Science Foundation Workshop on Next Generation Data Mining*, pp. 126-133, Baltimore, USA, November 2002

Dokas, P.; Ertoz, L.; Kumar, V.; Lazarevic, A.; Srivastava, J. & Tan, P.-N. (2002). Data mining for network intrusion detection, *Proceedings of NSF Workshop on Next Generation Data Mining*, pp. 21-30, November 2002

Domingos, P. & Richardson M. (2001). Mining the network value of customers, *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining*, pp. 57-66, San Francisco, USA, August 2001, ACM, California City

Field, A.; Hartel, P. & Mooij, W.(2001). Personal DJ, an Architecture for Personalised Content Delivery, *Proceedings of WWW10*, Hong Kong, China, May 2001, ACM 1-58113-348-0/01/0005

Garfinkel, S. (2001). Database Nation: The Death of the Privacy in the 21st Century, O'Reilly & Associates, Sebastopol, CA, USA, 2001.

Garofalakis, M..; Rastogi, R.; Seshadri, S. & Shim, K. (1999). Data Mining and the Web: Past, Present and Future, *Proceedings of 2nd ACM International Workshop on Web Information and Data Management (WIDM)*, pp. 43-47, Missouri, USA, November 1999, ACM, Kansas City

Gutiérrez, F.; Lerma, M.; Salgado, L. & Cantú, F. (2002). Biometrics and Data Mining: Comparison of Data Mining-Based Keystroke Dynamics Methods for Identity Verification, *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 221-245

Hernández, J.; Ochoa, A.; Andaverde, J. & Burlak. G. (2008). Biometrics in online assessments: A Study Case in High School Students, *Proceedings of 18th International Conference on Electronics, Communications and Computers Conielecomp 2008*, pp. 111-116

Hernández, J.; Ochoa, A.; Muñoz, J. & Burlak, G. (2006). Detecting cheats in online student assessments using Data Mining, *Proceedings of The 2006 International Conference on Data Mining (DMIN'2006)*, pp. 204-210, Las Vegas, USA, June 2006, Nevada City

Hill, W. & Terveen, L. (1996). Using Frequency-of-Mention in Public Conversations for Social Filtering, *Proceedings CSCW'96*, pp. 106-112, Boston, USA, November 1996, ACM, MA. City

ITEDU (2006). Biometrics. Consulted in http://et.wcu.edu/aidc/ August, 2006

Javitz, H. & Valdes, A. (1993). The NIDES Statistical Component: Description and Justification, Technical Report, Computer Science Laboratory, SRI International

Kumar, V.; Srivastava, J. & Lazarevic, A. (2005). Intrusion Detection: A survey, Managing Cyber Threats Issues, Approaches, and Challenges Chapter 2, Springer Verlag

Lauw, H.; Lim, E.; Tan, T. & Pang, H. (2005). Mining Social Network from Spatio-Temporal Events, *Proceedings of SIAM Data Mining Conference*, April 2005, Newport Beach

Lorrain F. & White H. (1971). structural equivalence of individuals in social networks, *In Journal of Mathematical Sociology*, pp. 49-80

Lundin, E.; Kvarnstrom, H. & Jonsson, E. (2002). A synthetic fraud data generation methodology, *In Lecture Notes in Computer Science ICICS 2002, Laboratories for Information Technology*, Singapore, December 2002, Springer Velag

Márquez, M.; Ojeda, S. & Hidalgo, H.(2008). Identification of behavior patterns in household solid waste generation in Mexicali's city: Study case, Journal of Resources, Conservation and Recycling, volumen 52, Issue 11, September 2008, pp. 1299–1306

Michigan Org (2007). Consulted on line at http://www.reachoutmichigan.org/funexperiments/agesubject/lessons/prints_ext.html, August 2007

Milgram S. (1967). *The small world problem*, Psychology Today, Vol., 2, pp. 60-67

Mobasher, B.; Dai H., Luo T.; Sun, Y. & Zhu J. (2000). Integrating web usage and content mining for more effective personalization, *Proceedings of Electronic Commerce and Web Technologies, First International Conference, EC-Web 2000*, pp. 165-176, ISBN 3-540-67981-2, London, UK , September 2000, Lecture Notes in Computer Science 1875 Springer 2000

MSU Michigan State University (2006). Quizzes and Exams: Cheating on exams and Quizzes Consulted in http://teachvu.vu.msu.edu/public/pedagogy/assessment/index.php?page_num=3, August 2006

Nielsen, J. & Loranger, H. (2006). Prioritizing Web Usability, *In* New Riders Press, , pp. 165-176, ISBN-13: 978-0-321-35031-2, Berkeley, CA

Ochoa, A. (2006). Más allá del Razonamiento Basado en Casos y una Aproximación al Modelado de Sociedades utilizando Minería de Datos, *Univ. Autónoma de Aguascalientes*, México, 1th Edition, Aguascalientes

Ochoa, A.; Agüero, M.; Mügerza, F.; Alvarado C.; Espino, M.; Jiménez, C.; Limón, J. & Valádez, M. (2007). Design and implementation of a Thematic Web Radio based on Social Data Mining and Cultural Algorithms, *Proceedings of CONTECSI'07*, Guanajuato, México, November 2007, León City

Ochoa, A.; Tcherassi, A.; Shingareva, I.; . Padméterakiris A.; Gyllenhaale J. & Hernández A. (2006). Italianitá: Discovering a Pygmalion effect on Italian communities using data

mining, *Proceedings of 7o Congreso de Computación CORE'2006*, ISSN: 1665-9899, México, México, May 2006, In Journal in Computing Science

Oliveira, S. & Zaiane, O. (2004). Toward Standardization in Privacy-Preserving Data Mining, *Proceedings of 3rd Workshop on Data Mining Standards, ACM SIGKDD*

Ponce, J.; Ochoa, A.; Pietsch, W. & Zolezzi-Hatsukimi, Z. (2007). Ahankara: Identify Bipolar Síndrome in User of Orkut with Data Mining, *Proceedings of ENC 2007*, Michoacan, Mexico, September 2007, IEEE, Morelia City

Ponce, J.; Ponce de León, E.; Padilla, F. Padilla, A. & Ochoa, A.(2006). Ant Colony Algorithm to solve Clique problem with Local Optimizer K-Opt (In Spanish), Journal Hífen, Urugaiana, Brazil, November 2006, pp. 191, ISSN 0103-1155

Rove, N. (2004). Cheating in Online Student Assessment: Beyond Plagiarism, *Online Journal of Distance Learning Administration*, Volume VII, Number II, Summer 2004 State University of West Georgia, Distance Education Center

Resnick, P.; Iacovou, N.; Suchak, M.; Bergstrom, P. & Riedl, J. (1994). GroupLens: An Open Architecture for Collaborative Filtering of Netnews, *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work CSCW'94*, pp. 175-186, Chapel Hill, NC, October 1994

Reynolds, R. (1998). An Introduction to Cultural Algorithms, *In Cultural Algorithms Repository*, http://www.cs.wayne.edu/~jcc/car.html

Rezgur, A.; Bouguettaya, A. & Eltoweissy. M. (2003). Privacy on the Web: Facts, Challenges, and Solutions, In IEEE Security & Privacy, pp. 40-49, November-December 2003

Schoeman, F. (1984). Philosophical Dimensions of Privacy, Cambridge University. Press

Schwartz, M. & Wood, D. (1993). Discovering shared interests using graph analysis., Communications, ACM, pp. 78–89

Varan, S. (2006). Crime Pattern Detection Using Data Mining, Oracle Corporation

Wahlstrom K., & Roddick J. (2000). On the Impact of Knowledge Discovery and Data Mining, *Proceedings of Australian Institute of Computer Ethics Conference (AiCE2000)*, Canberra, Australia, April 2000, Sydney City

Wasniowski, R. A. (2005). Using Data Fusion for biometric verification,*Transactions on Engineering Computing and Technology*, April 2005. pp. 72-74

Wasserman, S. & Faust, K. (1994). Social Network analysis: methods and applications, *Cambridge University Press*, ISBN-13: 9780521387071, Cambridge, UK

Wisher, R.; Curnov, C.; and Belanich, J. (2005). Verifying the Learner in distance learning, *Proceedings of 18 Annual Conference on Distance Teaching and Learning 2005*

Witten, I. & Frank E. (2005). *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, 2nd Edition, pp. 1-525, ISBN 0-12-088407-0, San Francisco

Xing, Dongshan; and Shen, Junyi (2004). Efficient data mining for web navigation patterns, *Journal of Information and Software Technology*. Volume 46, Issue 1, 1 January 2004, Pages 55-63

Xu, J.; Huang, Y. & Madey, G. (2003). A Research Support System Framework for Web Data Mining, *Proceedings of WSS'03: WI/IAT 2003 Workshop on Applications, Products of Web-based Support Systems*, ISBN 0-9734039-1-8, Halifax, Canada, Octuber 2003

Yao, J. & Yao Y. (2003). Web-based information retrieval support systems: building research tools for scientists in the new in-formation age, *Proceedings of the IEEE/WIC International Conference on Web Intelligence,* Halifax, Canada

Yao, Y. (2003). A framework for web-based research support systems, *Proceedings of Computer Software and Application Conference, COMPOSAC 2003,* Dallas, Texas

Zaiane, O.; Xin, M. & Han J. (1998). Discovering web access patterns and trends by applying OLAP and data mining technology on weblogs, In Advances in Digital Libraries, pp. 19-29

Zolezzi-Hatsukimi, Z. (2007). Implement social nets using Orkut, *Proceedings of CHI'07*, Nagoya, Japan

Data Mining and Knowledge Discovery in Real Life Applications

Edited by Julio Ponce and Adem Karahoca

ISBN 978-3-902613-53-0
Hard cover, 436 pages
Publisher I-Tech Education and Publishing
Published online 01, January, 2009
Published in print edition January, 2009

This book presents four different ways of theoretical and practical advances and applications of data mining in different promising areas like Industrialist, Biological, and Social. Twenty six chapters cover different special topics with proposed novel ideas. Each chapter gives an overview of the subjects and some of the chapters have cases with offered data mining solutions. We hope that this book will be a useful aid in showing a right way for the students, researchers and practitioners in their studies.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Julio Ponce, Alberto Hernández, Alberto Ochoa, Felipe Padilla, Alejandro Padilla, Francisco Álvarez and Eunice Ponce de León (2009). Data Mining in Web Applications, Data Mining and Knowledge Discovery in Real Life Applications, Julio Ponce and Adem Karahoca (Ed.), ISBN: 978-3-902613-53-0, InTech, Available from: http://www.intechopen.com/books/data_mining_and_knowledge_discovery_in_real_life_applications/data_mining_in_web_applications

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen